

目标极大似然估计的双稳健性及二阶偏差校正：一种半参数理论视角

李思默, 侯文

辽宁师范大学数学学院, 辽宁 大连

收稿日期: 2026年6月17日; 录用日期: 2026年7月11日; 发布日期: 2026年7月21日

摘要

目标极大似然估计(Targeted Maximum Likelihood Estimation, TMLE)是一种在半参数和非参数模型中构造局部有效估计量的通用框架。本文从半参数理论出发,系统阐述TMLE的核心理论基础。首先,针对平均处理效应(ATE)参数,严格推导其有效影响函数(EIF),并给出显式表达式。其次,深入分析TMLE的双重稳健性性质,通过偏差分解严格证明该性质成立的条件。作为重要参照,本文同时引入增强型逆概率加权估计量(AIPTW),通过并列对比两者的估计方程与偏差结构,揭示TMLE在双重稳健性实现路径上的独特优势。第三,利用von Mises展开揭示TMLE将一阶偏差降为二阶偏差的机制,阐明二阶余项乘积结构保证偏差的快速收敛。第四,专门讨论TMLE的正则性条件,包括Donsker条件的统计学含义、 $n^{-1/4}$ 收敛率要求的理论来源、交叉验证的绕行机制以及重叠性假设的影响。最后,证明TMLE的渐近正态性与半参数有效性。本文通过详细的数学推导,为研究者深入理解TMLE的半参数理论根基提供了系统性的技术基础。

关键词

目标极大似然估计, 有效影响函数, 双重稳健性, 平均处理效应, 半参数效率界

Double Robustness and Second-Order Bias Correction of Targeted Maximum Likelihood Estimation: A Semiparametric Theoretical Perspective

Simo Li, Wen Hou

School of Mathematics, Liaoning Normal University, Dalian Liaoning

Received: June 17, 2026; accepted: July 11, 2026; published: July 21, 2026

文章引用: 李思默, 侯文. 目标极大似然估计的双稳健性及二阶偏差校正: 一种半参数理论视角[J]. 应用数学进展, 2026, 15(7): 169-180. DOI: 10.12677/aam.2026.157313

Abstract

Targeted Maximum Likelihood Estimation (TMLE) is a general framework for constructing locally efficient estimators in semiparametric and nonparametric models. From a semiparametric perspective, this paper systematically presents the core theoretical foundations of TMLE. First, for the average treatment effect (ATE) parameter, we rigorously derive its efficient influence function (EIF) and provide an explicit expression. Second, we thoroughly analyze the double robustness property of TMLE and rigorously prove the conditions under which this property holds through bias decomposition. As an important reference, we simultaneously introduce the augmented inverse probability weighted estimator (AIPW) and, by juxtaposing their estimating equations and bias structures, reveal the unique advantages of TMLE in achieving double robustness. Third, we employ the von Mises expansion to demonstrate how TMLE reduces the first-order bias to second order, and we elucidate how the product structure of the second-order remainder ensures fast convergence of the bias. Fourth, we specifically discuss the regularity conditions for TMLE, including the statistical meaning of the Donsker condition, the theoretical origin of the $n^{-1/4}$ convergence rate requirement, the mechanism by which cross-fitting bypasses these conditions, and the impact of the overlap assumption. Finally, we prove the asymptotic normality and semiparametric efficiency of TMLE. Through detailed mathematical derivations, this paper provides a systematic technical foundation for researchers to gain a deep understanding of the semiparametric theoretical underpinnings of TMLE.

Keywords

Targeted Maximum Likelihood Estimation, Efficient Influence Function, Double Robustness, Average Treatment Effect, Semiparametric Efficiency Bound

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在因果推断中, 研究者常常面临一个核心挑战: 如何从观测数据中对某个因果参数进行无偏且高效地估计。传统的参数回归方法过度依赖于正确设定函数形式, 一旦模型设定出现偏差, 估计结果将产生难以忽视的一阶偏倚[1]。逆概率加权(IPTW)估计则依赖倾向评分模型的正确设定, 其偏差同样是一阶量级[2]。

为了突破这一困境, van der Laan 和 Rubin 于 2006 年提出了目标极大似然估计[3], 该方法在半参数模型中通过构造一个针对目标参数的波动子模型, 对初始密度估计进行更新, 使得最终的替代估计量能够求解有效影响函数估计方程, 从而在正则条件下获得局部有效的参数估计[4]-[6]。

TMLE 最引人注目的理论特性是其双重稳健性: 只要结局回归模型和治疗分配模型两者中有一个被正确指定, TMLE 对目标参数的估计就是一致的[7][8]。这一性质极大地增强了因果推断分析的可靠性[9][10]。关于 TMLE 的简单数值示例可参见文献[11]; 近期的理论进展进一步拓展了其适用性[12][13]。

本文旨在从纯理论视角系统梳理 TMLE 的核心理论基础。第 2 节设定模型与符号; 第 3 节推导平均处理效应的有效影响函数并给出 TMLE 的定义; 第 4 节提供主要理论结果的严格证明, 包括双重稳健性(含与 AIPW 的对比)、二阶偏差校正和渐近正态性; 第 5 节专门讨论 TMLE 的正则性条件; 第 6 节进行总结。

2. 模型的设定与符号

2.1. 数据结构与统计模型

令 $O_1, \dots, O_n$ 为从概率分布 $P_0 \in M$ 中独立同分布抽取的 n 个观测样本, 其中每个观测单元 $O = (W, A, Y)$ 的结构如下:

- $W \in W$: 基线协变量向量;
- $A \in \{0, 1\}$: 二元处理指示变量;
- $Y \in R$: 结果变量。

假设 M 由一个公共测度 μ 所支配, 那么每个概率测度 P 可以通过其密度 $p = dP/d\mu$ 来识别。设 P_n 为观测数据的经验分布, 即 $P_n f = \frac{1}{n} \sum_{i=1}^n f(O_i)$ 。

2.2. 目标参数与识别的模型的假设条件

定义潜在结果框架: 设 $Y(1)$ 和 $Y(0)$ 分别为个体在接受处理 $A=1$ 和 $A=0$ 时的潜在结果。平均处理效应(ATE)定义为:

$$\psi_0 = E[Y(1) - Y(0)].$$

为了从观测数据中识别 ATE, 通常需要以下可识别性条件:

- (1) 一致性: $Y = Y(1)A + Y(0)(1 - A)$;
- (2) 无混淆性: $\{Y(1), Y(0)\} \perp A | W$;
- (3) 重叠性: $0 < P(A=1|W) < 1$, 几乎必然成立。

在这些假设下, ATE 可表示为观测数据的泛函:

$$\psi(P) = E_W[E[Y|A=1, W] - E[Y|A=0, W]]. \quad (1)$$

记结局回归函数为:

$$\mu(a, w) = E[Y|A=a, W=w],$$

倾向性评分为:

$$g(w) = P(A=1|W=w).$$

目标参数 $\psi(P)$ 可视为从无限维概率分布 P 到实数空间 R 的一个光滑泛函(pathwise differentiable functional)。下面给出路径可微的严格定义。

定义 2.1 (光滑泛函) 设 $\Psi: M \rightarrow R$ 为统计泛函。若对任意 $P \in M$, 任意有界方向 h (满足 $\int h d\mu = 0$) 以及足够小的 δ , 导数

$$\left. \frac{\partial}{\partial \delta} \Psi(P + \delta h) \right|_{\delta=0}$$

存在且有界, 则称 Ψ 在 P 处是路径可微的。其导数可表示为某函数 $\phi(z; P)$ 的期望形式, 即:

$$\left. \frac{\partial}{\partial \delta} \Psi(P + \delta h) \right|_{\delta=0} = \int \phi(z; P) h(z) d\mu(z).$$

该函数 $\phi(z; P)$ 称为影响函数。在所有影响函数中, 方差最小的称为有效影响函数(EIF), 其方差即为半参数效率界。

3. 目标极大似然估计(TMLE)的理论框架

3.1. 有效影响函数(EIF)的推导

为了推导 ATE 的 EIF, 我们考虑一个参数路径 $P_\epsilon = P + \epsilon h$, 其中 h 满足 $\int h d\mu = 0$, 且 P_ϵ 为概率测度。ATE 泛函 $\Psi(P) = E_w[\mu(1, W) - \mu(0, W)]$ 依赖于条件期望 $\mu(A, W)$ 和边际分布 P_w 。我们需要分别计算对条件分布和边际分布的扰动。

首先, 记 $Q(A, W) = E[Y | A, W]$, 则 $\Psi(P) = \int [Q(1, w) - Q(0, w)] dP_w(w)$ 。沿着路径 P_ϵ , 泛函的导数为:

$$\begin{aligned} \left. \frac{\partial}{\partial \epsilon} \Psi(P_\epsilon) \right|_{\epsilon=0} &= \int \left[\left. \frac{\partial}{\partial \epsilon} Q_\epsilon(1, w) \right|_{\epsilon=0} - \left. \frac{\partial}{\partial \epsilon} Q_\epsilon(0, w) \right|_{\epsilon=0} \right] dP_w(w) \\ &\quad + \int [Q(1, w) - Q(0, w)] d \left(\left. \frac{\partial}{\partial \epsilon} P_{w, \epsilon} \right|_{\epsilon=0} \right)(w). \end{aligned}$$

根据半参数理论中的标准结果, 条件期望的路径导数与扰动函数的关系可表示为:

$$\left. \frac{\partial}{\partial \epsilon} Q_\epsilon(a, w) \right|_{\epsilon=0} = E \left[\frac{A}{g(W)} (Y - \mu(1, W)) \cdot h(O) | A = a, W = w \right].$$

实际上, 更直接的方法是使用影响函数的显式公式。已知对于 ATE 参数, 有效影响函数为:

$$\phi(z; P) = \frac{A}{g(W)} (Y - \mu(1, W)) - \frac{1-A}{1-g(W)} (Y - \mu(0, W)) + \mu(1, W) - \mu(0, W) - \Psi(P).$$

因此, 我们直接给出 ATE 的 EIF 显式表达式:

$$\phi(z; P) = \frac{A}{g(W)} (Y - \mu(1, W)) - \frac{1-A}{1-g(W)} (Y - \mu(0, W)) + \mu(1, W) - \mu(0, W) - \Psi(P). \quad (2)$$

该 EIF 具有两个关键性质:

- **均值为零:** $E_p[\phi(O; P)] = 0$ 。
- **正交性:** EIF 与干扰参数(μ 和 g)的得分函数正交, 这意味着在估计 μ 或 g 时产生的偏差在 EIF 中“解耦”, 为后续的双重稳健性提供了理论根基。

此外, EIF 的方差 $\text{Var}_p[\phi(O; P)]$ 给出了 $\Psi(P)$ 的半参数效率界, 即任何正则估计量的渐近方差的下界。

3.2. TMLE 的形式化定义与更新机制

TMLE 的核心思想是通过构造一个针对目标参数的一维波动子模型(fluctuation model), 对初始估计 μ_n 进行更新, 使得更新后的估计量 μ_n^* 满足 EIF 估计方程。具体构造步骤如下:

步骤 1 (初始估计): 使用数据自适应方法(如超级学习器 Super Learner)估计结局回归函数 $\mu_n(A, W)$ 和倾向性评分 $g_n(W)$ 。

步骤 2 (构造聪明协变量): 定义

$$H(A, W) = \frac{A}{g(W)} - \frac{1-A}{1-g(W)}.$$

注意到 $H(1, W) = 1/g(W)$, $H(0, W) = -1/(1-g(W))$ 。

步骤 3 (波动模型): 构造通过 μ_n 的参数子模型

$$\mu_{\epsilon}^*(A, W) = \text{expit}(\text{logit}(\mu_n(A, W)) + \epsilon \cdot H(A, W)),$$

其中 $\text{expit}(x) = 1/(1 + e^{-x})$, $\text{logit}(p) = \log(p/(1-p))$ 。该子模型在 $\epsilon = 0$ 处的得分函数恰好是 ϕ 中依赖于数据的第一部分。

步骤 4 (极大似然更新): 求解

$$\hat{\epsilon} = \arg \max_{\epsilon} \sum_{i=1}^n \log L(Y_i | A_i, W_i; \epsilon),$$

其中 L 是由 μ_{ϵ}^* 确定的似然函数(对于二元 Y 为伯努利似然, 对于连续 Y 为带 logistic 变换的高斯似然)。更新后的估计量为 $\mu_n^*(A, W) = \mu_{\hat{\epsilon}}^*(A, W)$ 。

步骤 5 (TMLE 估计量):

$$\hat{\psi}^{TMLE} = \frac{1}{n} \sum_{i=1}^n [\mu_n^*(1, W_i) - \mu_n^*(0, W_i)].$$

可以证明, 上述更新过程等价于求解 EIF 估计方程 $\frac{1}{n} \sum_i \phi(O_i; \hat{P}^*) = 0$, 其中 $\hat{P}^*$ 是更新后的概率分布估计。这一性质是 TMLE 具备良好渐近性质的根本原因。

4. 主要理论结果

4.1. 双重稳健性

定义 4.1 (双重稳健性) 一个估计量 $\hat{\psi}$ 称为双重稳健的, 如果下列两个条件中至少有一个成立时, $\hat{\psi} \rightarrow_p \psi_0$:

- (1) 结局回归模型 $\mu(a, w)$ 被正确指定(即 $\mu_n \rightarrow \mu_0$);
- (2) 处理分配模型 $g(w)$ 被正确指定(即 $g_n \rightarrow g_0$)。

4.1.1. AIPTW 估计量: 定义与性质

在深入分析 TMLE 的双重稳健性之前, 我们首先引入增强型逆概率加权估计量(AIPTW)作为重要的理论参照物。AIPTW 由 Robins 等[2] [7]提出, 其估计方程为:

$$\hat{\psi}^{AIPTW} = \frac{1}{n} \sum_{i=1}^n \left[\frac{A_i}{g_n(W_i)} (Y_i - \mu_n(1, W_i)) - \frac{1 - A_i}{1 - g_n(W_i)} (Y_i - \mu_n(0, W_i)) + \mu_n(1, W_i) - \mu_n(0, W_i) \right]. \quad (3)$$

AIPTW 的估计方程与 EIF(2)在结构上完全一致, 区别在于 AIPTW 直接将初始估计 μ_n 和 g_n 代入 EIF 的样本均值, 而不对 μ_n 进行任何更新。换言之, AIPTW 本质上是通过直接求解经验版本的 EIF 方程来构造估计量。

AIPTW 的双重稳健性可通过如下偏差分解看出。将 $Y = \mu_0(A, W) + \varepsilon$ 代入(3)并取期望, 得到:

$$\text{Bias}(\hat{\psi}^{AIPTW}) = E \left[\left(\frac{g_0(W)}{g_n(W)} - 1 \right) (\mu_0(1, W) - \mu_n(1, W)) \right] + E \left[\left(\frac{1 - g_0(W)}{1 - g_n(W)} - 1 \right) (\mu_n(0, W) - \mu_0(0, W)) \right].$$

该表达式的乘积结构($g_0 - g_n$ 与 $\mu_0 - \mu_n$ 的乘积)立即蕴含了双重稳健性: 当 $\mu_n \rightarrow \mu_0$ 或 $g_n \rightarrow g_0$ 时, 偏差趋于零。

4.1.2. TMLE 的双重稳健性证明

定理 4.2 (TMLE 的双重稳健性) 在正则条件下, TMLE 估计量 $\hat{\psi}^{TMLE}$ 是双重稳健的。

证明 设 μ_0 和 g_0 为真实的条件期望和倾向性评分, μ_n 和 g_n 为各自的估计。由 TMLE 的构造, 更新后的估

计量满足 EIF 估计方程:

$$P_n \phi(\cdot; \hat{P}^*) = 0,$$

其中 P_n 表示经验均值。展开该方程, 我们有:

$$\frac{1}{n} \sum_{i=1}^n \left[\frac{A_i}{g_n(W_i)} (Y_i - \mu_n^*(1, W_i)) - \frac{1-A_i}{1-g_n(W_i)} (Y_i - \mu_n^*(0, W_i)) + \mu_n^*(1, W_i) - \mu_n^*(0, W_i) - \hat{\psi}^{TMLE} \right] = 0.$$

整理可得:

$$\hat{\psi}^{TMLE} = \frac{1}{n} \sum_{i=1}^n \left[\frac{A_i}{g_n(W_i)} (Y_i - \mu_n^*(1, W_i)) - \frac{1-A_i}{1-g_n(W_i)} (Y_i - \mu_n^*(0, W_i)) \right] + \frac{1}{n} \sum_{i=1}^n [\mu_n^*(1, W_i) - \mu_n^*(0, W_i)].$$

注意到 μ_n^* 是 μ_n 的更新, 在正则条件下, μ_n^* 与 μ_n 的差异是 $o_p(1)$, 因此我们可以近似用 μ_n 代替 μ_n^* 以分析偏差。为简化, 我们直接分析初始估计对应的偏差, 但 TMLE 的更新保证了 EIF 方程成立, 因此偏差表达式可写为:

$$\hat{\psi}^{TMLE} - \psi_0 = P_n \left[\frac{A}{g_n(W)} (Y - \mu_n(1, W)) - \frac{1-A}{1-g_n(W)} (Y - \mu_n(0, W)) \right] + P_n [\mu_n(1, W) - \mu_n(0, W)] - \psi_0.$$

将 $Y = \mu_0(A, W) + \varepsilon$ 代入, 其中 $E[\varepsilon | A, W] = 0$, 得到:

$$\begin{aligned} \hat{\psi}^{TMLE} - \psi_0 &= P_n \left[\frac{A}{g_n(W)} (\mu_0(1, W) - \mu_n(1, W) + \varepsilon) - \frac{1-A}{1-g_n(W)} (\mu_0(0, W) - \mu_n(0, W) + \varepsilon) \right] \\ &\quad + P_n [\mu_n(1, W) - \mu_n(0, W)] - \psi_0. \end{aligned}$$

取期望(忽略经验过程的随机性, 因为 $n \rightarrow \infty$ 时 P_n 收敛到期望), 偏差的期望为:

$$\begin{aligned} Bias &= E \left[\frac{A}{g_n(W)} (\mu_0(1, W) - \mu_n(1, W)) - \frac{1-A}{1-g_n(W)} (\mu_0(0, W) - \mu_n(0, W)) \right] \\ &\quad + E [\mu_n(1, W) - \mu_n(0, W)] - \psi_0. \end{aligned}$$

由于 $\psi_0 = E[\mu_0(1, W) - \mu_0(0, W)]$, 上式可写为:

$$Bias = E \left[\left(\frac{A}{g_n(W)} - 1 \right) (\mu_0(1, W) - \mu_n(1, W)) \right] + E \left[\left(\frac{1-A}{1-g_n(W)} - 1 \right) (\mu_n(0, W) - \mu_0(0, W)) \right].$$

利用 $E[A|W] = g_0(W)$, 进一步化简:

$$Bias = E \left[\left(\frac{g_0(W)}{g_n(W)} - 1 \right) (\mu_0(1, W) - \mu_n(1, W)) \right] + E \left[\left(\frac{1-g_0(W)}{1-g_n(W)} - 1 \right) (\mu_n(0, W) - \mu_0(0, W)) \right].$$

即:

$$Bias = E \left[\frac{g_0(W) - g_n(W)}{g_n(W)} (\mu_0(1, W) - \mu_n(1, W)) \right] + E \left[\frac{g_n(W) - g_0(W)}{1-g_n(W)} (\mu_n(0, W) - \mu_0(0, W)) \right]. \quad (4)$$

现在, 如果 $\mu_n \rightarrow \mu_0$, 那么两个期望中的 $\mu_0 - \mu_n$ 和 $\mu_n - \mu_0$ 都趋于零, 因此偏差趋于零。如果 $g_n \rightarrow g_0$,

那么分子 $g_0 - g_n$ 和 $g_n - g_0$ 趋于零, 同时分母 g_n 和 $1 - g_n$ 有界(由重叠性假设), 因此整个偏差也趋于零。这就证明了双重稳健性: 只要 μ_n 或 g_n 中至少有一个是一致的, $\hat{\psi}^{TMLE}$ 就是一致的。

4.1.3. TMLE 与 AIPTW 的对比: 共性与差异

通过比较(3)与(4), 我们可以清晰地看到 TMLE 与 AIPTW 在双重稳健性上的共性: 两者的偏差结构均为两个近似误差的乘积形式, 因此只要 μ_n 或 g_n 中任意一个一致收敛, 偏差即趋于零。

然而, 两者在以下几个关键维度上存在本质差异:

(a) 估计方程的求解方式。 AIPTW 直接将初始估计 μ_n, g_n 代入 EIF 的样本均值, 得到一个“即插即用”的闭式估计量。而 TMLE 并不直接使用 μ_n , 而是通过波动模型将 μ_n 更新为 μ_n^* , 使得 $P_n \phi(\cdot; \hat{P}^*) = 0$ 精确成立。这一更新过程确保了 TMLE 的估计值始终位于参数空间内(例如, 对于二元结局, $\mu_n^* \in [0, 1]$), 而 AIPTW 的“替代估计量”可能超出参数空间, 在有限样本中产生不合理的估计值。

(b) 二阶偏差校正。 AIPTW 虽然双重稳健, 但其偏差量级仍为一阶的——具体而言, 其偏差由 $E[(\hat{g} - g_0)(\hat{\mu} - \mu_0)]$ 主导。TMLE 则将整体偏差量级降至二阶, 这是 TMLE 超越 AIPTW 的关键优势。

(c) 对 Donsker 条件的依赖。 AIPTW 的渐近正态性要求函数类 $\{\phi(\cdot; \hat{P})\}$ 满足 Donsker 条件, 否则经验过程项将不收敛。TMLE 在标准形式下同样依赖该条件, 但通过引入交叉验证(cross-fitting)可以绕过这一要求(详见第 5 节)。在这一点上, 两者在实践中的处理方式趋于一致。

(d) 有限样本稳定性。 TMLE 的似然化更新机制使其在有限样本中通常比 AIPTW 更为稳定, 其替代估计量(plug-in estimator)形式天然保证了其取值范围的合理性。

4.2. Von Mises 展开与二阶偏差

传统方法(如简单回归调整或 IPTW)的偏差是一阶的: $Bias = O_p(\mu_n - \mu_0)$ 或 $O_p(g_n - g_0)$ 。TMLE 的优势在于其偏差量级被降至二阶。为此, 本节首先建立 von Mises 展开的严格理论, 然后据此证明 TMLE 的二阶偏差性质。

4.2.1. Von Mises 展开的一般形式

von Mises 展开是泛函分析中泰勒展开在统计泛函上的推广[14] [15]。对于路径可微的统计泛函 $\Psi: M \rightarrow R$, 若它在 P_0 处是二阶路径可微的, 则对任意 $\hat{P} \in M$, 成立如下精确展开:

$$\Psi(\hat{P}) - \Psi(P_0) = \int \frac{\delta \Psi}{\delta P}(P_0)(d\hat{P} - dP_0) + \frac{1}{2} \iint \frac{\delta^2 \Psi}{\delta P^2}(P_0)(d\hat{P} - dP_0)^{\otimes 2} + R_3(\hat{P}, P_0),$$

其中 $\frac{\delta \Psi}{\delta P}(P_0)$ 是影响函数, R_3 是三阶余项。

在 TMLE 的渐近理论中, 通常使用一阶 von Mises 展开, 其形式为:

$$\Psi(\hat{P}) - \Psi(P_0) = P_n \phi(\cdot; P_0) - P_n \phi(\cdot; \hat{P}) + (P_n - P_0) [\phi(\cdot; \hat{P}) - \phi(\cdot; P_0)] + R_2(\hat{P}, P_0). \quad (5)$$

这里 P_n 是经验测度, $R_2(\hat{P}, P_0)$ 是二阶余项。该展开可以通过添加和减去 $P_n \phi(\cdot; \hat{P})$ 并利用

$\Psi(\hat{P}) - \Psi(P_0) = -P_0 \phi(\cdot; \hat{P}) + (P_n - P_0) [\phi(\cdot; \hat{P}) - \phi(\cdot; P_0)]$ 等恒等式得到。

引理 4.3 (von Mises 展开的标准形式) 设 $\Psi: M \rightarrow R$ 是路径可微的, $\hat{P}, P_0 \in M$ 。假设 P_0 关于 $\hat{P}$ 绝对连续, 且 Radon-Nikodym 导数 $dP_0/d\hat{P}$ 在 $\hat{P}$ 下平方可积。则展开(5)成立, 且二阶余项 $R_2(\hat{P}, P_0)$ 满足:

$$R_2(\hat{P}, P_0) = O_p(\|\hat{P} - P_0\|^2),$$

其中 $\|\cdot\|$ 是适当的范数(例如 Hellinger 距离)。对于 ATE 参数, 具体地

$$R_2(\hat{P}, P_0) = O_p(\|\mu_n - \mu_0\| \cdot \|g_n - g_0\|).$$

4.2.2. TMLE 中 Von Mises 展开的特殊化

将引理应用于 TMLE, $\hat{P}$ 取为 TMLE 更新后的分布估计 $\hat{P}^*$ 。TMLE 的核心目标正是通过波动模型进行更新, 使得替代偏倚项 $P_n\phi(\cdot; \hat{P}^*)$ 为零。事实上, TMLE 的极大似然更新等价于求解 EIF 估计方程 $\sum_i \phi(O_i; \hat{P}^*) = 0$, 即 $P_n\phi(\cdot; \hat{P}^*) = 0$ 。代入展开式(3)得到:

$$\Psi(\hat{P}^*) - \Psi(P_0) = P_n\phi(\cdot; P_0) + (P_n - P_0)[\phi(\cdot; \hat{P}^*) - \phi(\cdot; P_0)] + R_2(\hat{P}^*, P_0). \quad (6)$$

在正则条件下, 经验过程项 $(P_n - P_0)[\phi(\cdot; \hat{P}^*) - \phi(\cdot; P_0)] = o_p(n^{-1/2})$, 只要 $\phi(\cdot; \hat{P}^*)$ 和 $\phi(\cdot; P_0)$ 之间的 $L^2(P_0)$ 距离收敛到零, 并且函数类 $\{\phi(\cdot; \hat{P}^*)\}$ 满足 Donsker 条件。因此, TMLE 估计量的渐近行为由下式主导:

$$\Psi(\hat{P}^*) - \Psi(P_0) = P_n\phi(\cdot; P_0) + o_p(n^{-1/2}) + R_2(\hat{P}^*, P_0).$$

这就是 TMLE 实现二阶偏差校正的理论依据——当 $R_2 = o_p(n^{-1/2})$ 时,

$$\sqrt{n}(\Psi(\hat{P}^*) - \Psi(P_0)) \rightarrow_d N(0, P_0[\phi^2]).$$

4.2.3. 二阶余项的具体展开与正则性条件

对 ATE 泛函应用 von Mises 展开, 可以得到二阶余项的具体形式。假设 μ_n 和 g_n 分别为 μ_0 和 g_0 的估计。经过复杂的变分计算(见[4]), 得到:

$$\begin{aligned} R_2(\hat{P}^*, P_0) = & E_{P_0} \left[\frac{g_n(W) - g_0(W)}{g_n(W)} (\mu_n(1, W) - \mu_0(1, W)) \right] \\ & - E_{P_0} \left[\frac{g_n(W) - g_0(W)}{1 - g_n(W)} (\mu_n(0, W) - \mu_0(0, W)) \right]. \end{aligned} \quad (7)$$

该表达式清晰地显示了一个乘积结构: 它由 $\mu_n - \mu_0$ 和 $g_n - g_0$ 的乘积的期望组成。因此,

$$R_2(\hat{P}^*, P_0) = O_p(\|\mu_n - \mu_0\| \cdot \|g_n - g_0\|),$$

其中 $\|\cdot\|$ 表示 $L^2(P_0)$ 范数。

为保证 von Mises 展开的渐近性质成立, TMLE 渐近理论所需的详细正则性条件(包括一致性、Donsker 条件、二阶余项收敛速度等)将在第 5 节集中讨论。

4.2.4. 二阶偏差定理

现在可以正式陈述并证明 TMLE 的二阶偏差性质。

定理 4.4 (偏差的二阶性质) 在正则条件下, TMLE 的剩余偏差为:

$$\text{Bias}(\hat{\psi}^{TMLE}) = O_p(\|\mu_n - \mu_0\| \cdot \|g_n - g_0\|). \quad (8)$$

证明 由 von Mises 展开(4)及 TMLE 更新后满足 $P_n\phi(\cdot; \hat{P}^*) = 0$, 可得:

$$\hat{\psi}^{TMLE} - \psi_0 = P_n\phi(\cdot; P_0) + (P_n - P_0)[\phi(\cdot; \hat{P}^*) - \phi(\cdot; P_0)] + R_2(\hat{P}^*, P_0).$$

取期望, 由于 $E[P_n\phi(\cdot; P_0)] = 0$, 且经验过程项在正则条件下期望为零(实际上是 $o_p(n^{-1/2})$), 期望也是小量,

我们得到:

$$\text{Bias}(\hat{\psi}^{\text{TMLE}}) = E[R_2(\hat{P}^*, P_0)] + o(1).$$

根据(7)的乘积结构, 存在常数 C 使得

$$|R_2(\hat{P}^*, P_0)| \leq C \cdot \mu_n - \mu_0 \cdot g_n - g_0$$

依概率成立。因此,

$$\text{Bias}(\hat{\psi}^{\text{TMLE}}) = O_p(\mu_n - \mu_0 \cdot g_n - g_0).$$

若 μ_n 和 g_n 均以速率至少 $n^{-1/4}$ 收敛到其真值(例如, 通过使用交叉验证和适当的学习器实现), 则

$$R_2 = O_p(n^{-1/2}).$$

因此, TMLE 是 $\sqrt{n}$ -一致的, 即使 μ_n 和 g_n 各自都无法以 $n^{-1/2}$ 的速率收敛。这正是 TMLE 区别于传统估计量的核心优势: 它将估计任务转化为同时学习两个函数的问题, 而只需各自以 $n^{-1/4}$ 的速率收敛即可达到 $n^{-1/2}$ 的总体估计精度。

4.3. 渐近正态性与有效性

4.3.1. 渐近线性估计量及其影响函数

为了理解 TMLE 的渐近分布, 首先回顾渐近线性估计量的概念。

定义 4.5 (渐近线性估计量) 估计量 $\hat{\psi}_n$ 称为在 P_0 处是渐近线性的, 若存在一个实值函数 $\tilde{\phi}(O)$ (称为影响函数, 满足 $E_{P_0}[\tilde{\phi}(O)] = 0$ 且 $E_{P_0}[\tilde{\phi}(O)^2] < \infty$), 使得:

$$\hat{\psi}_n - \psi_0 = \frac{1}{n} \sum_{i=1}^n \tilde{\phi}(O_i) + o_p(n^{-1/2}).$$

对于渐近线性估计量, 由中心极限定理直接可得:

$$\sqrt{n}(\hat{\psi}_n - \psi_0) \rightarrow_d N(0, \text{Var}_{P_0}[\tilde{\phi}(O)]).$$

因此, 确定估计量的影响函数是推导渐近方差的核心步骤。

4.3.2. TMLE 的影响函数

由 von Mises 展开(4)及 TMLE 更新后 $P_n \phi(\cdot; \hat{P}^*) = 0$, 我们有:

$$\Psi(\hat{P}^*) - \Psi(P_0) = P_n \phi(\cdot; P_0) + (P_n - P_0)[\phi(\cdot; \hat{P}^*) - \phi(\cdot; P_0)] + R_2(\hat{P}^*, P_0).$$

记 $\hat{\psi}^{\text{TMLE}} = \Psi(\hat{P}^*)$, $\psi_0 = \Psi(P_0)$ 。在以下两个正则条件下:

(1) 经验过程项可忽略: $(P_n - P_0)[\phi(\cdot; \hat{P}^*) - \phi(\cdot; P_0)] = o_p(n^{-1/2})$ (由 Donsker 条件保证);

(2) 二阶余项可忽略: $R_2(\hat{P}^*, P_0) = o_p(n^{-1/2})$ (由定理 4.4 中乘积结构的收敛速度保证),

得到:

$$\hat{\psi}^{\text{TMLE}} - \psi_0 = \frac{1}{n} \sum_{i=1}^n \phi(O_i; P_0) + o_p(n^{-1/2}).$$

这意味着 TMLE 是渐近线性的, 其影响函数恰好是 EIF $\phi(O_i; P_0)$ 。

4.3.3. 渐近正态分布与方差

定理 4.6 (渐近正态性与局部有效性) 在正则条件下, 当 $R_2(\hat{P}^*, P_0) = o_p(n^{-1/2})$ 且 Donsker 条件成立时, TMLE 估计量 $\hat{\psi}^{TMLE}$ 满足:

$$\sqrt{n}(\hat{\psi}^{TMLE} - \psi_0) \rightarrow_d N(0, \sigma_{EIF}^2), \quad (9)$$

其中 $\sigma_{EIF}^2 = \text{Var}_{P_0}[\phi(O; P_0)]$ 为半参数效率界。此时, TMLE 是局部有效的: 当 μ_0 和 g_0 均被一致估计时, TMLE 的渐近方差达到半参数模型下所有正则估计量方差的下界。

证明 由 TMLE 的渐近线性表示:

$$\hat{\psi}^{TMLE} - \psi_0 = \frac{1}{n} \sum_{i=1}^n \phi(O_i; P_0) + o_p(n^{-1/2}),$$

两边乘以 $\sqrt{n}$:

$$\sqrt{n}(\hat{\psi}^{TMLE} - \psi_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(O_i; P_0) + o_p(1).$$

由于 $E_{P_0}[\phi(O; P_0)] = 0$ 且 $\text{Var}_{P_0}[\phi(O; P_0)] = \sigma_{EIF}^2$, 由中心极限定理:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \phi(O_i; P_0) \rightarrow_d N(0, \sigma_{EIF}^2).$$

结合 Slutsky 定理, 即得(9)。

效率论证: 在半参数估计理论中, 由卷积定理(convolution theorem)可知: 对于任意正则估计量, 其渐近方差至少等于有效影响函数的方差——这是所有正则估计量能够达到的最小方差下界, 称为半参数效率界。具体地,

$$\lim_{n \rightarrow \infty} \text{Var}_{P_0}[\sqrt{n}\hat{\psi}_n] \geq \text{Var}_{P_0}[\phi(O; P_0)].$$

TMLE 由于在渐近展开中恰好以 EIF 作为影响函数, 其渐近方差达到了这一下界, 因此是渐近有效的。当 μ_0 和 g_0 均被正确估计时, TMLE 的渐近方差等于半参数效率界, 故称其为局部有效。

4.3.4. 方差估计量的构造

在实际应用中, EIF 的真值 $\phi(O; P_0)$ 是未知的, 但可以用其估计量 $\phi(O; \hat{P}^*)$ 来估计方差。

定理 4.7 (方差估计量的相合性) 在正则条件下, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \phi(O_i; \hat{P}^*)^2$ 是 σ_{EIF}^2 的相合估计, 即

$$\hat{\sigma}^2 \rightarrow_p \sigma_{EIF}^2.$$

证明 由于 $\hat{P}^* \rightarrow_p P_0$ 且 $\phi(\cdot; P)$ 在 P_0 处是连续的(在适当的度量下), 由连续映射定理可得 $\phi(O; \hat{P}^*) \rightarrow_p \phi(O; P_0)$ 。再结合大数定律,

$$\frac{1}{n} \sum_{i=1}^n \phi(O_i; \hat{P}^*)^2 \rightarrow_p E_{P_0}[\phi(O; P_0)^2] = \sigma_{EIF}^2.$$

基于此, TMLE 的渐近置信区间可构造为:

$$\hat{\psi}^{TMLE} \pm z_{1-\alpha/2} \cdot \frac{\hat{\sigma}}{\sqrt{n}},$$

其中 $z_{1-\alpha/2}$ 是标准正态分布的 $1-\alpha/2$ 分位数。这正是推导渐近正态分布方差的实际意义——为统计推断提供理论基础。

5. 正则性条件：Donsker 条件、收敛率与重叠性

TMLE 的渐近性质(双重稳健性、二阶偏差、渐近正态性)依赖于若干正则性条件。本节系统讨论这些条件的含义、来源及实践中的处理策略。

5.1. Donsker 条件：含义与作用

TMLE 的渐近线性表示要求经验过程项可忽略：

$$(P_n - P_0) \left[\phi(\cdot; \hat{P}^*) - \phi(\cdot; P_0) \right] = o_p(n^{-1/2}). \quad (10)$$

这通常通过要求函数类 $F = \{\phi(\cdot; P) : P \in M\}$ 是 P_0 -Donsker 类来保证。Donsker 条件本质上是限制函数类的复杂度。

5.2. $n^{-1/4}$ 收敛率要求的来源

二阶余项(7)的乘积结构 $R_2 = O_p(\|\mu_n - \mu_0\| \cdot \|g_n - g_0\|)$ 决定了收敛率要求。设 $\|\mu_n - \mu_0\| = O_p(n^{-\alpha})$ ， $\|g_n - g_0\| = O_p(n^{-\beta})$ ，则 $R_2 = O_p(n^{-(\alpha+\beta)})$ 。要使 $R_2 = o_p(n^{-1/2})$ ，只需 $\alpha + \beta > 1/2$ 。特别地， $\alpha = \beta = 1/4$ 即满足。因此 TMLE 不要求任一函数单独达到 $n^{-1/2}$ 速率，而是利用乘积结构的“折衷”效应，这使得 TMLE 适用于收敛率较慢的现代机器学习方法。

5.3. 交叉验证(Cross-Fitting)：绕行 Donsker 条件

交叉验证(样本拆分)可有效绕行 Donsker 条件[8][13]。标准流程为：将样本随机分成 K 折，对每一折 k ，使用训练集估计 μ_n^{-k}, g_n^{-k} ，在验证集上执行 TMLE 更新，得到 $\hat{\psi}_k^{TMLE}$ ，最终估计为

$\hat{\psi}^{CV-TMLE} = \frac{1}{K} \sum_{k=1}^K \hat{\psi}_k^{TMLE}$ 。其核心机制是：每一折中，估计量在训练集上训练，与验证集独立，从而经验过程的条件期望为零，无需全局 Donsker 条件。

5.4. 重叠性(Positivity)假设及其影响

重叠性 $0 < g(W) < 1$ 几乎必然成立，其作用包括重叠性保证权重有界、控制二阶余项量级；有限样本中 $\hat{g}_n$ 接近 0 或 1 时会导致权重爆炸。可对倾向性评分进行截断缓解。

表 1 汇总了主要正则性条件及其应对措施。

Table 1. Main regularity conditions for the asymptotic theory of TMLE

表 1. TMLE 渐近理论的主要正则性条件

条件	表述	主要作用	实践中如何满足/绕行
重叠性	$0 < g(W) < 1$ 几乎必然	保证权重有界，控制偏差与方差	截断、C-TMLE、稳定估计
一致性	$\ \hat{P}^* - P_0\ \rightarrow_p 0$	渐近展开的前提	使用强一致性学习器(如 SuperLearner)
Donsker 条件	$\{\phi(\cdot; P)\}$ 是 P_0 -Donsker 类	经验过程项 $o_p(n^{-1/2})$	使用交叉验证(CV-TMLE)绕行
二阶余项速率	$\ \mu_n - \mu_0\ \cdot \ g_n - g_0\ = o_p(n^{-1/2})$	保证 $\sqrt{n}$ 一致性	使用收敛率至少 $n^{-1/4}$ 的估计方法
光滑性	$\phi(\cdot; P)$ 在 P_0 处连续	方差估计相合性	由泛函可微性自动满足

6. 结语

本文从半参数理论的视角，系统阐述了目标极大似然估计的核心理论基础。主要贡献如下：

(1) 有效影响函数与效率: TMLE 通过构造使用估计量求解 EIF 估计方程, 实现了半参数意义上的局部有效性。

(2) 双重稳健性: 只要结局模型或处理分配模型至少有一个被正确指定, TMLE 即提供一致估计。双重稳健性的核心数学结构源于 EIF 方程的性质: 当 $\int \phi(z; \hat{P}) dP_n = 0$ 时, 代入后天然消去了来源于 μ 或 g 的单方向偏差。

(3) 二阶偏差: 利用 von Mises 展开, TMLE 将偏差从一阶降为二阶。这意味着 μ 和 g 只需以 $n^{-1/4}$ 的速率收敛, TMLE 即可达到 $\sqrt{n}$ 一致, 这使其在数据自适应学习中具有显著优势。

(4) 似然化框架: TMLE 的更新过程是一个完整的统计推断流程: 从初始密度估计出发, 沿得分方向调整, 再做替代。这一框架相比直接求解估计方程的方法, 在参数范围保持、有限样本稳定性及复杂参数扩展方面具有优势。

TMLE 自 2006 年提出以来, 已在因果推断、流行病学、生物统计等领域获得了广泛应用, 并发展为包含交叉验证 TMLE、协作 TMLE 等扩充形式的完整 “Targeted Learning” 理论体系。对于希望从事半参数估计理论或应用研究的工作者而言, 深入理解 TMLE 的上述理论根基, 不仅是方法学上的必要训练, 更是开展高质量因果推断研究的重要基础。

基金项目

本研究由 2022 年度辽宁省研究生教育教学改革研究项目(2022-180-39510165)资助。

参考文献

- [1] Tsiatis, A.A. (2006) Semiparametric Theory and Missing Data. Springer.
- [2] Robins, J.M. and Rotnitzky, A. (1992) Recovery of Information and Adjustment for Dependent Censoring Using Surrogate Markers. In: *AIDS Epidemiology*, Birkhäuser Boston, 297-331. https://doi.org/10.1007/978-1-4757-1229-2_14
- [3] van der Laan, M.J. and Rubin, D. (2006) Targeted Maximum Likelihood Learning. *The International Journal of Biostatistics*, 2, Article 11. <https://doi.org/10.2202/1557-4679.1043>
- [4] van der Laan, M.J. and Rose, S. (2011) Targeted Learning: Causal Inference for Observational and Experimental Data. Springer.
- [5] van der Laan, M.J. and Gruber, S. (2010) A Targeted Maximum Likelihood Estimator of a Causal Effect on a Bounded Continuous Outcome. *The International Journal of Biostatistics*, 6, Article 26. <https://doi.org/10.2202/1557-4679.1260>
- [6] Gruber, S. and van der Laan, M.J. (2009) Targeted Maximum Likelihood Estimation: A Gentle Introduction. U.C. Berkeley Division of Biostatistics Working Paper Series, Paper 252.
- [7] Bang, H. and Robins, J.M. (2005) Doubly Robust Estimation in Missing Data and Causal Inference Models. *Biometrics*, 61, 962-973. <https://doi.org/10.1111/j.1541-0420.2005.00377.x>
- [8] Zheng, W. and van der Laan, M. J. (2010) Asymptotic Theory for Cross-Validated Targeted Maximum Likelihood Estimation. U.C. Berkeley Division of Biostatistics Working Paper Series, Paper 273.
- [9] Moccia, C., Moirano, G., Popovic, M., Pizzi, C., Fariselli, P., Richiardi, L., et al. (2024) Machine Learning in Causal Inference for Epidemiology. *European Journal of Epidemiology*, 39, 1097-1108. <https://doi.org/10.1007/s10654-024-01173-x>.
- [10] Ross, R.K., Montoya, L.M., Goin, D.E., Díaz, I. and Renson, A. (2025) Constructing Targeted Minimum Loss/Maximum Likelihood Estimators: A Simple Illustration to Build Intuition. *American Journal of Epidemiology*, 195, 1818-1821. <https://doi.org/10.1093/aje/kwaf261>
- [11] Rosenblum, M. and van der Laan, M. J. (2010) Simple Examples of Estimating Causal Effects Using Targeted Maximum Likelihood Estimation. U.C. Berkeley Division of Biostatistics Working Paper Series, Paper 262.
- [12] Kim, M. (2026) Efficient Targeted Maximum Likelihood Estimation of Average Treatment Effects under Structured Outcome Models with Unknown Error Distributions.
- [13] Xu, Y. and van der Laan, M.J. (2026) Adaptive Targeted Maximum Likelihood Estimation of the Mean Potential Outcome under a Treatment Rule.
- [14] Fernholz, L.T. (1983) von Mises Calculus for Statistical Functionals. Springer-Verlag.
- [15] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1998) Efficient and Adaptive Estimation for Semiparametric Models. Springer.